

Effect of resampling schemes on significance analysis of clustering and ranking

Atieh Mirshahvalad* and Martin Rosvall†
Integrated Science Lab, Department of Physics, Umeå University, Sweden

Olivier H. Beauchesne‡ and Éric Archambault§
Science-Metrix, Montréal, Canada.

Community detection helps us simplify the complex configuration of networks, but communities are reliable only if they are statistically significant. To detect statistically significant communities, one approach is to repeatedly perturb the original network and analyze the communities. But the perturbation approach is reliable only if we understand how the results depend on the underlying assumptions of the perturbation method. Here we explore how maintaining link correlations in resampling schemes affects the significance of communities in citation networks. We compare maintained link correlations in non-parametric article resampling with parametric resampling of citations that reduce link correlations in multinomial and Poisson resampling. While multinomial resampling maintains the variance of individual link weights and eliminates correlations between connected links, Poisson resampling eliminates any link correlations. For significance analysis of ranking and clustering, we find that it is more important to capture the variance of individual link weights than the correlations between link weights. We also find that Poisson resampling underestimates the variance of link weights. Therefore, when only link weights are available and neither article resampling nor multinomial resampling is possible, we suggest a simple parametric resampling scheme that generates link variances close to link variances of non-parametric article resampling. Nevertheless, when we highlight and summarize important structural changes in science, we find that the more link correlations we maintain in the resampling scheme, the earlier we can predict structural change.

1. INTRODUCTION

Researchers use network theory[1] to better understand complex systems[2–5] that have many interacting components[6–10]. In network theory, there is a great interest in detecting communities[11–21]; that is, tightly interconnected structural patterns of the network. Community detection helps us simplify the structure of the network because the communities often correspond to functional units of the system, but communities are reliable only if they are statistically significant [22–25]. Detecting statistically significant communities is possible when we have many instances of the network because then we can iteratively cluster the instances to find a set of community descriptions. From these descriptions, we can assess the significance of each community. But most often, we only have a single observation of the real network. To overcome this challenge and detect significant communities of real networks, we need a stochastically sound procedure that generates resamples of the single raw network. A common approach for generating resamples of the raw network is to use perturbation techniques [26–29], in which researchers assume a model that introduces random noise to the data. The idea behind the perturbation approach is fairly simple: a network is the result of accumulating some natural events with specific properties. By using a perturbation method, researchers make an assumption about the events that generate the observed network. Based on the underlying assumption, they

can imitate the process of network generation and create various samples of the raw network. Afterwards, they can analyze recreated networks and aggregate the result of the corresponding communities such that they can determine which communities of the raw network are significant and to which degree. But when using the perturbation approach, we must ask: How much do the results of significance analysis depend on the perturbation method and its underlying assumption? More specifically, how important are the link correlations in the resampling scheme?

Testing and analyzing the perturbation method is possible only if we have more information available about the raw network. Here we aim to explore how much the link correlations of the resampling schemes affect the results of significance analysis in the case of weighted, directed citation networks. In weighted networks, the significance of communities not only depend on the strength of the links, but also on the correlations between links: which links are simultaneously strong. In previous work, and with data limited to citation counts between journals, we used Poisson resampling without link correlations to generate bootstrap networks[28]. That is, independently from other links, we resampled the weight of each weighted directed link from a Poisson distribution with mean equal to the original link weight. With access to article level data, we now can accurately resample articles to better assess the significance of communities under the condition that link correlations are fully preserved. In addition, we can better understand the effects of eliminated link correlations in Poisson resampling. Our dataset includes citations between more than 900,000 scientific articles published in about 11,000 journals in the years 1984-2010. For a specific year, we can build a citation network between scientific journals in which each link weight between two journals A and B represents the number of times that articles published in journal A cite articles pub-

*Electronic address: atieh.mirshahvalad@physics.umu.se

†Electronic address: martin.rosvall@physics.umu.se

‡Electronic address: olivier.hbeauchesne@science-metrix.com

§Electronic address: eric.archambault@science-metrix.com

lished in journal B .

With the article level citation data, we can explore how link correlations in different resampling schemes affect the statistical properties of communities. But because communities are aggregated measures of a network, in order to better understand the role of the resampling scheme on the statistical analysis of communities, we will investigate how much resampling schemes affect the finest detail levels: individual links. We also will investigate what role the link correlations of the resampling scheme play on the properties of neighbor links: links that share a common source journal. The link variance and link correlations contribute to the statistical analysis of aggregated measures of a network such as clustering or ranking.

2. METHODS

To understand which effect the link correlations of the resampling scheme have on assessing significant communities, we compare a resampling scheme with full preserved link correlations (article resampling), a resampling scheme with half preserved link correlations (multinomial resampling) and a resampling scheme with no preserved link correlation (Poisson resampling), see Fig. 1. Keeping half of the link correlations in multinomial resampling corresponds to conserving within-link correlations and destroying between-link correlations.

In the first scenario, if the article level citation is available, the scientific process that eventually results in the citation network is either an article published or not. We can actually resample from real articles and assemble the bootstrap networks by aggregating articles in journals. The process of article resampling is simple: assume that we have a pool of all articles that participate in our citation network. We pick an article from this pool and add the citations of this article from the journal in which the article was published to the cited journals. Then we put this article back in the pool. We continue this process until we reach the same number of articles that the original network has. Since one article might cite articles from different journals, the event of article resampling, automatically introduces correlation between the links of the bootstrap networks (Fig. 1(a)). For example, suppose that an article x that was published in *PNAS* cited an article in *Nature* and also an article in *Science*. When article x is chosen in the resampling process, it increases the link weight from *PNAS* to *Nature* and the link weight from *PNAS* to *Science* concurrently. In addition, because an article might cite articles of a specific journal more than once, article resampling introduces correlations within links (Fig. 1(a)). To investigate how these correlations affect the significance analysis, we compare article resampling with multinomial resampling, which keeps correlations within links but destroys correlations between links.

In the second scenario, we consider a situation in which the full article level citation is not available but the detailed information about the events that generate a specific link is known. This case naturally suggests resampling every link weight based on a multinomial resampling. To generate boot-

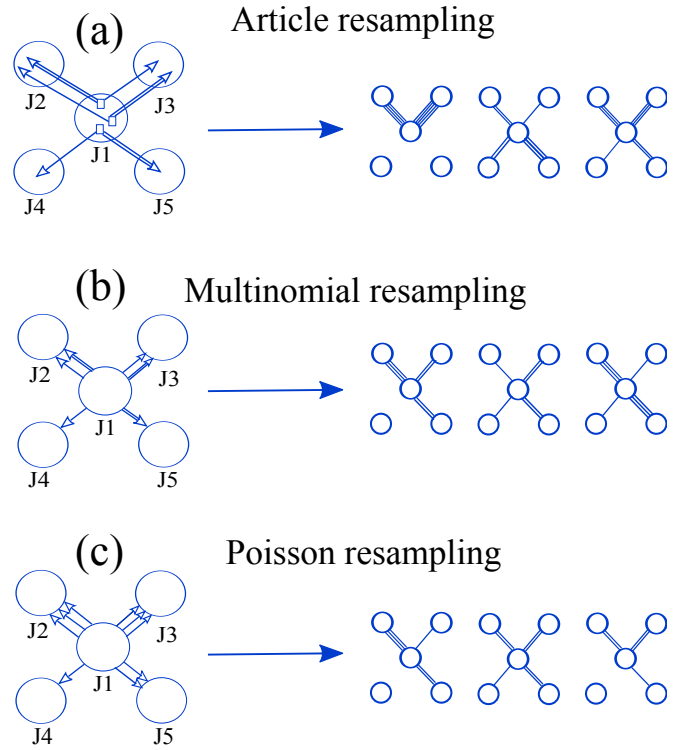


FIG. 1: **Link correlation preservation in different resampling schemes** (a) Article resampling preserves correlations between links and also correlations within links. For example, an article in journal J1 might cite articles from journal J2 together with articles in journal J3 (correlations between links). An article in journal J1 might also cite another journal J2 more than once (correlations within links). The right hand side shows some examples of possible resampled networks that necessarily keep correlation between and within links. (b) Multinomial resampling only preserves the correlations within links. The examples of resampled networks on the right hand side shows that they could be generated without keeping between-links correlations. (c) Poisson resampling does not preserve any link correlation. Every link of a resampled networks is generated independently of others.

strap networks, we resample every link weight of the raw network from a multinomial distribution, which is estimated based on article level data information, and we do this independently for all links of the raw network. Multinomial resampling does not preserve correlations between links, but it does preserve the correlations within links (Fig. 1(b)). As a result, multinomial resampling creates an intermediate stage between a completely destroyed link correlation scheme (Poisson resampling) and a fully preserved link correlation scheme (article resampling). The question is: how much is the result of significance analysis in the reduced-link correlation scheme (multinomial resampling) different from the fully preserved-link correlation scheme (article resampling). In section 3, we will show that significant clusters in multinomial resampling are close to significant clusters in article resampling, which implies that the role of link dependency on significance analysis of clusters is small.

In the third scenario, if we only have access to the jour-

nal citation data, we can assume that the scientific process that eventually results in the citation network is either a researcher adds a citation between two journals or not. This event is equivalent to assuming that citations are spread according to a Poisson distribution. In such a case, the event of citation between a pair of journals is independent of citation between another pair. The process of Poisson resampling for generating bootstrap networks is as follows: we resample every link weight of the raw network from a Poisson distribution with mean equal to the original link weight, and we do this independently for all links of the raw network. Poisson resampling not only automatically ignores the correlation between links but it also ignores the correlations within links: Poisson resampling assumes independent events that participate in making a specific link (Fig. 1(c)). The question is: how much do the results of significance analysis specifically depend on the independent link weights assumption? We aim to investigate which role these correlations play in the bootstrap networks and what effect these differences would have on the significance analysis of the clusters.

3. RESULTS AND DISCUSSION

In order to investigate the effect of link correlations on the significance analysis of clusters, we create 1000 bootstrap networks based on a resampling scheme. Then we search for significant clusters, cluster cores, which we define as the biggest subset of nodes in each cluster that gathered together in more than 90% of the bootstrap networks. For clustering, we use *infomap*, which is an information theoretic algorithm that reveals regularities in a given network based on how information flows on that network [30]. Figure 2 shows the difference between significant cluster cores of article, multinomial, and Poisson resampling in terms of *normalized information distance*. Normalized information distance is defined as:

$$d_{\max} = 1 - \frac{I(C, C')}{\max(H(C), H(C'))} \quad (1)$$

where $H(\dots)$ refers to Shannon entropy and $I(C, C')$ is the mutual information between the significant cores of the two resampling schemes that tells us how similar they are. Mutual information between two clusters C and C' is described as:

$$I(C; C') = \sum_{c, c'} P(c, c') \log \frac{P(c, c')}{P(c)P(c')} \quad (2)$$

where $P(c, c')$ is the joint probability distribution between two clusterings c and c' . $P(c)$ and $P(c')$ refer to the marginal probability distributions. The second term in Eq. 1 is called normalized mutual information.

If C and C' are identical, then the normalized mutual information is equal to 1, which means that, by knowing one cluster structure, we know the other one, and conversely, if C and C' are completely independent, by knowing one, we learn nothing about the other one and the normalized mutual information between them would be 0. We use normalized information distance for comparing clusterings because it is a sound

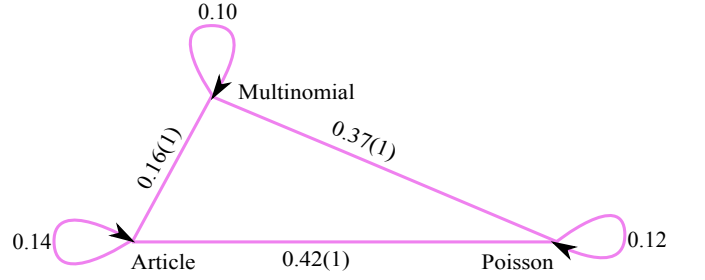


FIG. 2: **The differences between significant clusters' cores in different resampling schemes.** We calculate normalized information distance (d_{\max}) between the significant cores of the two corresponding methods with respect to the page rank. All values correspond to an average over at least 2000 runs.

metric[31]. Figure 2 shows that the difference between significant cores of article and multinomial resampling is in the same order as the difference between each of these schemes with itself, and both of them are considerably different from Poisson resampling. Although multinomial resampling does not hold the correlation between citations and article resampling does, our results show that between-link dependency does not have a great impact on the significance analysis of clusters.

We further explore the effects of link variance in a concrete example. Figure 3 shows the alluvial diagram of the three resampling schemes over years 1989-1993. In the alluvial diagram, each block represents a specific module in a given year. The height of a block represents its importance in terms of PageRank [32]. In a block, the lighter colors correspond to the insignificant part of the module; the bigger this area is, the more insignificant nodes that module has. The white vertical gap between blocks determines separate modules, and the numbers under each block correspond to the year. As shown in the figure, all three resampling schemes agree on the separation of *Nuclear & Particle physics* from *General physics* as an independent stand-alone module in 1993. In fact, given the data we have, we see the separation of Nuclear & particle physics as a significant stand-alone module in 1993, but by no means we conclude that this is the emergence of a new field. While Nuclear & particle physics was considered a research area long before 1993, it takes some time before it shows up in the structure of the network. The separation of a module in citation network data reflects the specialization of scientific fields according to the journal's citation pattern. But as Fig. 3 shows, different schemes consider the separation of the Nuclear & particle physics module differently in time. For example, in article resampling, this module is a completely insignificant part of General physics even in 1989, while in Poisson and multinomial resampling, this happens later. So the process of becoming insignificant could provide us a signal about important changes that might happen in the future; apparently, article resampling can give this signal sooner than multinomial resampling, and multinomial resampling can give it sooner than Poisson resampling. In summary, for significance analysis of communities, within-link correlations play a more major role than between-link correlations. Conserving link correlations in a resampling scheme helps us to recog-

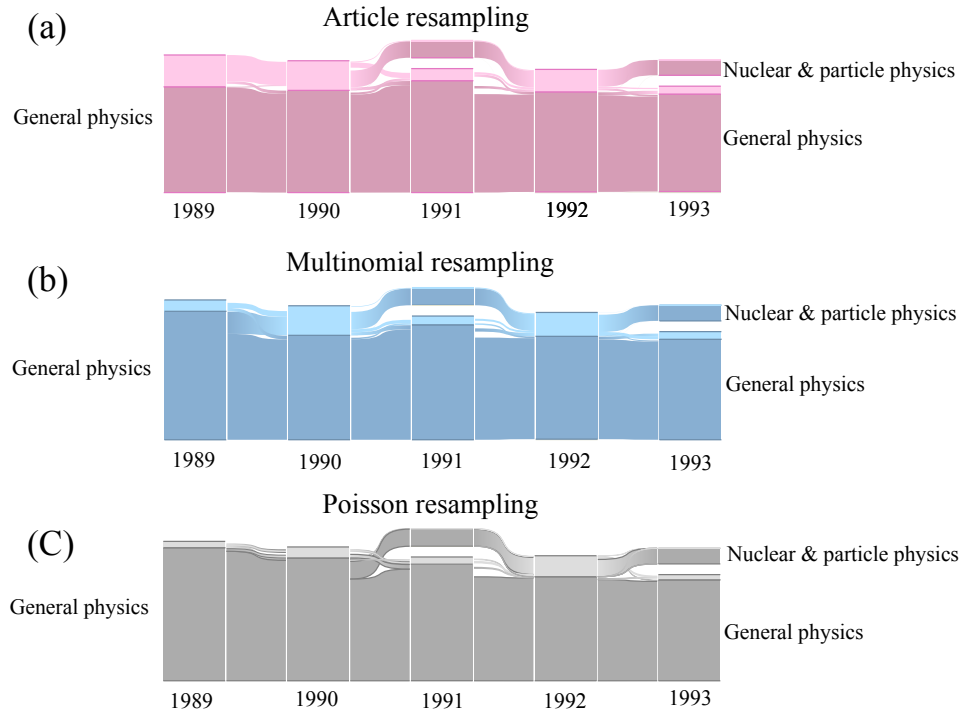


FIG. 3: **The separation of Nuclear and particle physics from the physics module.** While all three resampling schemes agree on the separation of *Nuclear & particle physics* from *General physics* into an independent stand-alone module by 1993, article resampling emits a signal about this change sooner than multinomial or Poisson resampling.

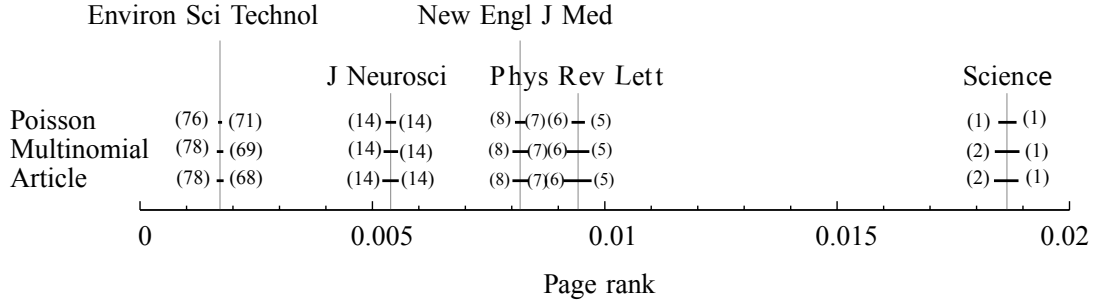


FIG. 4: **The variation of the PageRank for top rank journals based on different resampling schemes.** In agreement with the result of single link variance analysis, our analysis shows that core structures in article and multinomial resampling are much more similar to each other than in the Poisson resampling. The article resampling is the biggest perturbation, in which the 95% confidence interval for the PageRank is broader than in multinomial or article resampling. Multinomial and article resampling were second and third, respectively.

nize the changes of a network earlier. In another words, as we neglect link correlations in our resampling, we postpone the emergence of insignificant clusters further. Consequently, we postpone a signal that could give us a hint about changes that might happen in the future.

We also investigate PageRank as another aggregated measure to test and analyze the effect of a resampling scheme and its correlation strategy on the network. In calculating PageRank, the importance of a node (a journal in our citation network) corresponds to the importance of nodes that cite this node. So the full network indirectly participates in calculating the PageRank of a node. Figure 4 shows how much the PageRank of some top journals would vary based on the resampling

scheme. The length of each line accords with an interval that covers the variation of PageRank for a given journal in a given resampling scheme. The numbers on the left/right hand side of each line correspond to the minimum/maximum rank order of each journal for a resampling scheme. *Science* has the largest PageRank value in the raw network, and so it is the first journal in the rank order. In the Poisson resampling, *Science* always maintains its first rank position in the ranking list. But in multinomial and article resampling, sometimes the rank of *Science* decreases to the second position. In a similar fashion, the rank order of PRL (*Physical Review Letters*), NEJM (*New England Journal of Medicine*) and J Neurosci (*Journal of Neuroscience*) is changed based on the resampling scheme

that is used. In general, the PageRank of a node varies more in article resampling compared to Poisson or multinomial resampling. In this respect, we study the effect of resampling schemes on the rank order of all nodes in the network. We sample pairs of nodes (i, j) from the rank order that we obtain from a resampling scheme and compare them with the rank order that we obtain from another resampling scheme. We sample pairs of nodes proportional to their PageRank and measure the similarity between the two rank orders in terms of normalized mutual information. If for all possible pairs in the two-rank order, the node with the highest rank in one order is the same in the other order, the mutual information between the two rank orders would be one. The more different the two rank orders are (less common pair orders), the smaller the MI between them would be. If the two rank orders do not have any common pair orders, the MI between them would be zero. In a quantitative analysis of the rank order for the different resampling schemes, we find that the normalized information distance (Eq. 1) between two different rankings that are generated with the same resampling scheme, on average, is about 26 percent larger for article resampling than for Poisson resampling and 23 percent larger for multinomial resampling than for Poisson resampling. For ranking, article resampling has the biggest variation, but multinomial resampling without correlations between links varies almost as much as article resampling. Multinomial resampling can explain almost all ranking variances of article resampling with correlations between links.

The link dependency strategy of a resampling scheme does not have a great impact on ranking (Fig. 4) or on the significance analysis of clusters (Fig. 2). The reason it doesn't have a great impact is manifested in the behavior of resampling schemes at the detailed link level.

To better understand the results of resampling schemes on aggregated measures (ranking and clustering), we investigate what role the link correlations of a resampling scheme play on the properties of neighbor links. We investigate whether a preserved link correlation scheme, e.g. article resampling, has an expressive effect on the correlation between link weights or not. Then we look at the variation of individual links on each resampling scheme. Our results show that the multinomial resampling completely matches with the article resampling on the link level. Where the probabilities of different link weights are unknown for the purpose of performing multinomial resampling, we propose a simplistic model that can estimate the probabilities.

3.1. Pairwise link variance

Article resampling introduces dependencies between link weights: an article can cite papers in different journals, so picking that article adds citations to more than one journal simultaneously. Here we want to measure how much these neighbor links are correlated in the resampled networks. We call two links *neighbor links* if they share a common source node. Figure 5 shows that in article resampling, for most cases, neighbor links are not correlated, but in some cases,

there is a slight correlation between neighbor links. To check if this is a significant correlation or not, we compare article resampling with its close acquaintance which has independent assumptions between links: multinomial resampling. Figure 5 shows that when we compare article and multinomial resampling, the correlations between neighbor links are small. In fact, we could say that most neighbor links are not correlated. Those neighbor links that are slightly correlated tend to be correlated rather than anti-correlated. As we saw in the beginning of section 3, this slight correlation doesn't have a great impact on the significant cluster cores or ranking of nodes. As shown, the dependency between links has a small effect on significant cluster cores, but nevertheless it influences the time that insignificant clusters emerge and can give a clue about important changes that might happen in the future.

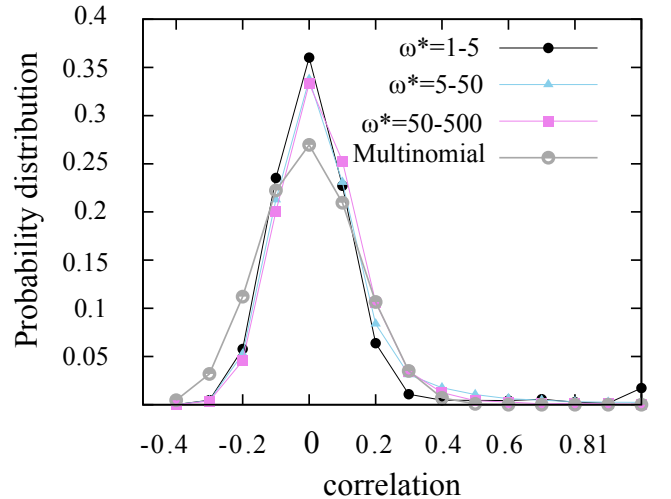


FIG. 5: **Article resampling does not strongly preserve the dependency of neighbor links.** The correlation distribution for a pair of neighbor links where at least one of them has a specific weight. By definition, article resampling introduces correlation to the neighboring links and multinomial resampling ignores any correlation. By comparing, we see that the result of correlation distribution confirms that most correlations of article resampling are not significant, when we compare them with the multinomial resampling as null mode. All points correspond to an average of at least 50 runs.

3.2. Single link variance

Figure 6 shows how much a specific link weight, w^* , varies based on the resampling scheme. When the link weight is very small, for example, $w^* = 1$, we see that the probability distribution of link weights in Poisson resampling perfectly matches with the article resampling (Fig. 6(a)). The link weight equal to one means that only one article contributes to the citation between two journals, so the chance of picking that article is $\frac{1}{N}$, where N is the total number of articles. Therefore, after resampling N articles, the chance of getting that specific paper k times is $\frac{1}{N}^k \binom{N-1}{N-k}$, which, in the limit of large N , coincides with the definition of *Poisson*(1). But

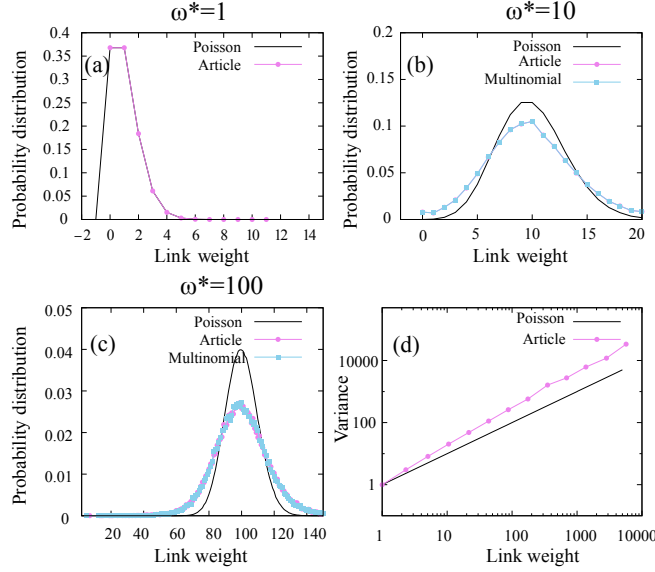


FIG. 6: Comparing the probability distribution of link weights in article resampling with Poisson resampling and multinomial resampling (a) For low link weight ($w^* = 1$), the Poisson resampling precisely coincides with article resampling. (b,c) For medium values of link weight ($w^* = 10$) and high values of link weight ($w^* = 100$), the Poisson resampling underestimates the article resampling. The variance of the distribution in article resampling is much higher than in Poisson resampling. For example, for $w^* = 10$, the variance of article resampling is $\sigma_{art}^2 = 19$, while the variance of Poisson resampling is $\sigma_{poiss}^2 = 10$. Similarly, for link weight $w^* = 100$, the variance of article resampling is $\sigma_{art}^2 = 290$, while the variance of Poisson resampling is $\sigma_{poiss}^2 = 100$. The variance of multinomial resampling is quite close to the article resampling, which confirms that the multinomial model imitates article resampling and make the distribution broader than Poisson resampling. (d) The variance of link weights in article resampling and Poisson resampling averaged over all resamples. All points correspond to averaging over 1000 runs.

when the link weight between two journals is higher than one, for example, medium values like $w^* = 10$ in Fig. 6(b) or high values like $w^* = 100$ in Fig. 6(c), we see that the variance of Poisson resampling underestimates the variance of article resampling. This happens because citations can come in groups: for a link where its weight w^* is medium/high, there are A articles ($A \leq w^*$) that contribute to that weight, and sometimes articles might add more than one citation. So, although article resampling gives the same average weight as Poisson resampling, the variance of that weight in article resampling would be higher than for Poisson resampling. In summary, high weights cause greater differences between the variance of article resampling and Poisson resampling (Fig. 6(d)).

Indeed, although Poisson resampling assumes an enormous amount of binomial events that produce a specific link weight, article resampling tells us that the observed link weight is the outcome of multinomial events. In multinomial resampling, every link weight is generated from a multinomial distribution independently from other links. Although multinomial resampling assumes independency between links' weights and article resampling does not, Fig. 6(b,c) shows that multinomial resampling completely matches article resampling on the link level. Multinomial resampling intrinsically considers group citations, and therefore it can generate higher variance than Poisson resampling. But what if the probabilities of different link weights are unknown for a given network? To estimate the probabilities, we look at the number of papers that contribute to a link with a specific weight. Figure 7(a) shows

that when link weight w is high, the number of papers that contribute to generating that link weight $N_P(w)$ is far from the value of the weight itself. Figure 7(b) shows that, when the link weight increases, the fraction of single citations that contribute to making that weight is reduced. To generate higher link weights, more group citations (2,3,... citations) contribute. As Fig. 7(a) shows, the number of papers that contribute to generating a link weight w scales as $w^{0.9}$ for all years. We use this information to build a model for estimating the multinomial distribution when the probabilities of different link weights in a given network are not known. For each weight w , we assume that it is generated out of papers with only one or two citations. We can simply estimate the number of papers with one citation N_1 and the number of contributing papers with two citations N_2 by solving the following linear equation system:

$$\left. \begin{aligned} N_P(w) &= N_1 + N_2 = w^{0.9} \\ N_1 + 2N_2 &= w \end{aligned} \right\} \Rightarrow \begin{aligned} N_1 &= 2w^{0.9} - w \\ N_2 &= w - w^{0.9} \end{aligned} \quad (3)$$

After estimating N_1 and N_2 , we suggest resampling every link weight by using the following minimal model:

$$Poisson(N_1) + 2Poisson(N_2) \quad (4)$$

The variance that we could get out from this model is:

$$\begin{aligned} Var(Poisson(N_1) + 2Poisson(N_2)) &= \\ 3w - 2w^{0.9} \end{aligned} \quad (5)$$

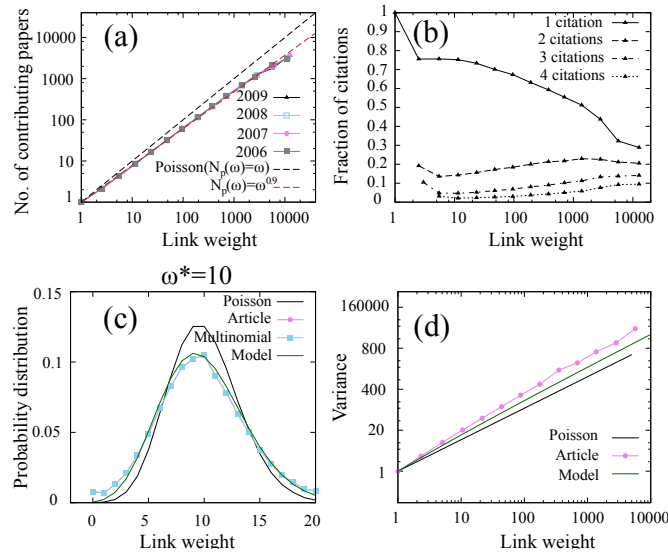


FIG. 7: **The high variance of article and multinomial resampling can be estimated by a simple model that extends Poisson resampling to account for papers that contribute with multiple citations to the same journal.** (a) Average number of papers that contribute to a specific link weight in logarithmic scale. For all years, the number of papers with weight w ($N_p(w)$) fits to the function x^α with exponent $\alpha = 0.9$. (b) The fraction of 1, 2, 3 and 4 citations that contribute to building a specific link weight w . Compared to low link weights, high link weights have a lower fraction of papers with only one citation and a higher fraction of papers with 2, 3 or 4 citations. (c) The probability distribution of link weight $w^* = 10$ for 4 cases: Poisson resampling, article resampling, multinomial resampling, and the minimal model. The high variance of article/multinomial resampling could be estimated by the model. (d) The model can generate higher variance than Poisson resampling for different link weights. However, it could not generate exactly as high a variance as article resampling.

In Fig. 7(c), we show the probability distribution of link weight $w^* = 10$ for four cases: Poisson resampling, article resampling, multinomial resampling, and the proposed model. As shown, the high variance of article/multinomial resampling could be estimated by the minimal model. However, this estimation is not exact because the minimal model does not take into account group citations with three citations or more. In summary, the model can generate higher variance than Poisson resampling for different link weights, but it can not generate exactly as high a variance as article resampling (Fig. 7(d)).

4. CONCLUSION

Link correlation of a resampling scheme influences the significance analysis of communities and ranking. We compare three scenarios: full preserved link correlations (article resampling), half preserved link correlations (multinomial re-

sampling), and no preserved link correlations (Poisson resampling). We found that the result of significance analysis in multinomial resampling that only conserves correlation within individual links almost matches with article resampling that conserves correlations both within and between links. We conclude that the role of correlation within individual links is greater than the role of correlation between links. Nevertheless, we found that conserving link correlation in a resampling scheme can provide an early hint of possible changes of the network in the future. These findings can help researchers to better understand and assess reliable significant communities and structural changes for a given network.

Acknowledgments

We are grateful to Sara de Luna and Deborah Kolp for many valuable discussions. MR was supported by Swedish Research Council grant 2009-5344.

[1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010)
[2] A. Vespignani, *Nat Phys* **8**, 32 (2012)
[3] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. L. Barabási, *Nature* **407**, 651 (2000)
[4] J. Kleinberg, *Nature* **406**, 845 (2000)
[5] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002)
[6] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)

[7] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003)
[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006)
[9] M. Sales-Pardo, R. Guimer, A. A. Moreira, and L. A. N. Amaral, *Proc. Natl. Acad. Sci. USA* **104**, 15224 (2007)
[10] A. Clauset, C. Moore, and M. E. J. Newman, *Nature* **453**, 98 (2008)
[11] M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **99**, 7821 (2002)

- [12] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, *Proc. Natl. Acad. Sci. USA* **101**, 2658 (2004)
- [13] M. E. J. Newman, *Phys. Rev. E* **69**, 066133 (2004)
- [14] L. Danon, A. Díaz-Guilera, and A. Arenas, *Stat. Mech.* **2006**, P11010 (2006)
- [15] V. B. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Stat. Mech.* **2008**, P10008 (2008)
- [16] M. B. Hastings, *Phys. Rev. E* **74**, 035102 (2006)
- [17] M. Rosvall and C. T. Bergstrom, *Proc. Natl. Acad. Sci. USA* **104**, 7327 (2007)
- [18] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Nature* **435**, 814 (2005)
- [19] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, *Nature* **466**, 761 (2010)
- [20] M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **103** (2006)
- [21] S. Fortunato, *Physics Reports* **486**, 75 (2010)
- [22] V. Spirin and L. A. Mirny, *Proc. Natl. Acad. Sci. USA* **100**, 12123 (2003)
- [23] Y. Hu, Y. Nie, H. Yang, J. Cheng, Y. Fan, and Z. Di, *Phys. Rev. E* **82**, 066106 (2010)
- [24] A. Lancichinetti, F. Radicchi, and J. Ramasco, *Phys. Rev. E* **81**, 046110 (2010)
- [25] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, *PLoS ONE* **6**, e18961 (2011)
- [26] D. Gfeller, J.-C. Chappelier, and P. De Los Rios, *Phys. Rev. E* **72**, 056135 (2005)
- [27] B. Karrer, E. Levina, and M. E. J. Newman, *Phys. Rev. E* **77**, 046119 (2008)
- [28] M. Rosvall and C. T. Bergstrom, *PLoS ONE* **5**, e8694 (2010)
- [29] A. Mirshahvalad, J. Lindholm, M. Derln, and M. Rosvall, *PLoS ONE* **7**, e33721 (2012)
- [30] M. Rosvall and C. T. Bergstrom, *Proc. Natl. Acad. Sci. USA* **105**, 1118 (2008)
- [31] N. X. Vinh, K. Epps, and J. Bailey, *Mach. Learn. Res.* **11**, 2837 (2010)
- [32] S. Brin and L. Page, *Computer Networks and ISDN Systems* **30**, 107 (1998)